

# IEE 578: Regression Analysis Final Project Report

## Regression Analysis of World Happiness Score 2018

Submitted to Dr. Douglas C. Montgomery on December 4, 2019

Submitted By Varun Agrawal (1215318065)

# **Table of Contents**

1.	Introduction	5
2.	Description of Data	5
,	2.1 Missing value	5
3.	Analysis	6
	3.1 Response with each regressor	6
	3.1.1 Happiness Score vs GDP per capita	6
	3.1.2 Happiness Score vs Social Support	7
	3.1.3 Happiness Score vs Health Life Expectancy	9
	3.1.4 Happiness Score vs Freedom	10
	3.1.5 Happiness Score vs Generosity	11
	3.1.6 Happiness Score vs Corruption	13
	3.2 Model Analysis	14
	3.2.1 Check for Multicollinearity	14
	3.2.2 Full-Model Analysis.	15
	3.2.3 Reduced-Model Analysis	18
•	3.3 Model Adequacy Checking	21
	3.4 Model Validation	22
4.	Conclusion	24
5.	References	25

# **List of Tables**

Table 1 Summary Statistics (Score vs GDP)	6
Table 2 Summary of Fit (Score vs GDP)	7
Table 3 Analysis of Variance (Score vs GDP)	7
Table 4 Parameter Estimate (Score vs GDP)	7
Table 5 Summary Statistics (Score vs Social Support)	8
Table 6 Summary of Fit (Score vs Social Support)	8
Table 7 Analysis of Variance (Score vs Social Support)	8
Table 8 Parameter Estimates (Score vs Social Support)	8
Table 9 Summary Statistics (Score vs Health).	9
Table 10 Summary of Fit (Score vs Health)	9
Table 11 Analysis of Variance (Score vs Health)	10
Table 12 Parameter Estimates (Score vs Health)	
Table 13 Summary Statistics (Score vs Freedom)	10
Table 14 Summary of Fit (Score vs Freedom)	
Table 15 Analysis of Variance (Score vs Freedom).	11
Table 16 Parameter Estimates (Score vs Freedom).	11
Table 17 Summary Statistics (Score vs Generosity)	12
Table 18 Summary of Fit (Score vs Generosity)	12
Table 19 Analysis of Variance (Score vs Generosity)	
Table 20 Parameter Estimates (Score vs Generosity)	
Table 21 Summary Statistics (Score vs Corruption)	13
Table 22 Summary of Fit (Score vs Corruption)	
Table 23 Analysis of Variance (Score vs Corruption)	14
Table 24 Parameter Estimates (Score vs Corruption)	14
Table 25 Parameter Estimates (Multicollinearity)	14
Table 26 Effect Summary (Full-Model Analysis).	
Table 27 Summary of Fit (Full-Model Analysis)	16
Table 28 Analysis of Variance (Full-Model Analysis)	
Table 29 Parameter Estimates (Full-Model Analysis)	
Table 30 Effect Summary (Reduced-Model Analysis)	
Table 31 Summary of Fit (Reduced-Model Analysis)	
Table 32 Analysis of Variance (Reduced-Model Analysis)	
Table 33 Summary Statistics (Model Adequacy)	
Table 34 Summary Statistics (Model Validation)	23

# **List of Figures**

Figure 1 Happiness Score vs GDP per capita (Scatter Plot)	6
Figure 2 Happiness Score vs Social Support (Scatter Plot)	7
Figure 3 Happiness Score vs Health Life Expectancy (Scatter Plot)	9
Figure 4 Happiness Score vs Freedom (Scatter Plot)	10
Figure 5 Happiness Score vs Generosity (Scatter Plot)	11
Figure 6 Happiness Score vs Corruption (Scatter Plot)	13
Figure 7 Actual by Predicted Plot (Full-Model)	15
Figure 8 Residual by Predicted Plot (Full-Model)	16
Figure 9 Studentized Residual Plot (Full-Model)	16
Figure 10 Prediction Profiler (Full-Model)	17
Figure 11 Normal Plot (Full-Model)	17
Figure 12 Actual by Predicted Plot (Reduced-Model)	19
Figure 13 Residual by Predicted Plot (Reduced-Model)	20
Figure 14 Prediction Profiler (Reduced-Model)	20
Figure 15 Normal Plot (Reduced-Model)	20
Figure 16 Studentized Residuals (Reduced-Model)	
Figure 17 Residual Happiness Score (Model Adequacy)	22
Figure 18 Residual Plot (Model Validation)	23

#### 1. Introduction

This report is a landmark survey of the state of global happiness for the year 2018 that ranks 156 countries by how happy their citizens perceive themselves to be. It focuses on how happiness has evolved over the past dozen years, with a focus on the technologies, social norms, conflicts, and government policies that have driven those changes. Leading experts describe how measurements of well-being can be used effectively to assess the progress of nations.

For the regression analysis, the response variable is the **Happiness Score**. The regressors (attributes) or the predictor variables are –

- (1) Economy or GDP per capita
- (2) Social Support
- (3) Healthy Life Expectancy
- (4) Freedom
- (5) Generosity
- (6) Corruption

The multilinear regression analysis is performed on the data and to do the model validation, the data set for the year 2019 is used and validated from the prediction equation received from the fitted model of the analysis.

## 2. Description of Data

The data set consists of 156 rows and seven columns in which the happiness score corresponds to the response column and the rest six are the regressor columns. The source of the data is – (View data)

## 2.1 Missing value

There is one value that is missing in the data set that corresponds to the corruption for row 21 (with the happiness score of 6.774). There are various ways to take care for this missing value such as calculating parameters associated with measure of central tendency like mean, median, or mode for that column and replace that missing value with one of them or the deletion of the entire row and analyzing the remaining data or look the other sources to get that data.

In this case, the mean of the remaining 155 values for the corruption was taken and that value was filled in place of the missing value.

## 3. Analysis

The entire regression analysis was performed in the JMP Pro 14 software. The data set was exported from <u>Excel</u> to JMP. The following procedure is followed to perform the analysis:

### 3.1 Response with each regressor

The response (happiness score) was analyzed with each regressor to see their significance in the further regression analysis. Following results are obtained for each attribute –

### 3.1.1 Happiness Score vs GDP per capita

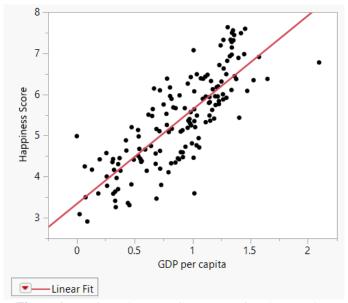


Figure 1 Happiness Score vs GDP per capita (Scatter Plot)

 Table 1 Summary Statistics (Score vs GDP)

Summary Statistics							
	٧	alue	Low	er 95%	Upper 95%	Signif. Prob	
Correlation	0.80	2124	0.	738006	0.851887	<.0001*	
Covariance	0.35	1938					
Count		156					
Variable		N	lean	Std De	ev		
GDP per cap	oita	0.89	1449	0.39192	21		
Happiness S	core	5.37	5917	1.11950	06		

#### Linear Fit

Happiness Score = 3.3333994 + 2.2912336\*GDP per capita

 Table 2 Summary of Fit (Score vs GDP)

Summary of Fit					
RSquare	0.643403				
RSquare Adj	0.641087				
Root Mean Square Error	0.670688				
Mean of Response	5.375917				
Observations (or Sum Wgts)	156				

## Lack of Fit

**Table 3** Analysis of Variance (Score vs GDP)

Analysis of Variance						
Source	DF	Sum of Squares	Mean Square	F Ratio		
Model	1	124.98774	124.988	277.8598		
Error	154	69.27273	0.450	Prob > F		
C. Total	155	194.26047		<.0001*		

Table 4 Parameter Estimate (Score vs GDP)

Parameter Estimates						
Term	Estimate	Std Error	t Ratio	Prob> t		
Intercept	3.3333994	0.133783	24.92	<.0001*		
GDP per capita	2.2912336	0.137454	16.67	<.0001*		

## 3.1.2 Happiness Score vs Social Support

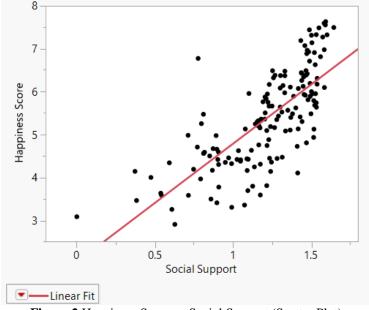


Figure 2 Happiness Score vs Social Support (Scatter Plot)

 Table 5 Summary Statistics (Score vs Social Support)

Summary Statistics							
	٧	alue	Low	er 95%	Upper 95%	Signif. Prob	
Correlation	0.7	4576	0.	666756	0.80819	<.0001*	
Covariance	0.25	2445					
Count		156					
Variable		N	/lean	Std De	ev		
Social Supp	ort	1.21	3237	0.30237	72		
Happiness S	core	5.37	5917	1.11950	06		

## Linear Fit

Happiness Score = 2.0260319 + 2.7611129\*Social Support

Table 6 Summary of Fit (Score vs Social Support)

Summary of Fit	
RSquare	0.556158
RSquare Adj	0.553276
Root Mean Square Error	0.748249
Mean of Response	5.375917
Observations (or Sum Wgts)	156

### Lack of Fit

Table 7 Analysis of Variance (Score vs Social Support)

Analysis of Variance							
Source	DF	Sum of Squares	Mean Square	F Ratio			
Model	1	108.03956	108.040	192.9705			
Error	154	86.22091	0.560	Prob > F			
C. Total	155	194.26047		<.0001*			

Table 8 Parameter Estimates (Score vs Social Support)

Parameter Estimates							
Term	Estimate	Std Error	t Ratio	Prob> t			
Intercept	2.0260319	0.248478	8.15	<.0001*			
Social Support	2.7611129	0.198764	13.89	<.0001*			

## 3.1.3 Happiness Score vs Health Life Expectancy

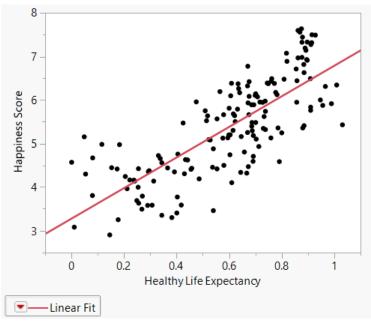


Figure 3 Happiness Score vs Health Life Expectancy (Scatter Plot)

Table 9 Summary Statistics (Score vs Health)

Summary Statistics							
	Value	Lower 95%	Upper 9	5%	Signif. Prob		
Correlation	0.775814	0.704568	0.8315	75	<.0001*		
Covariance	0.215029						
Count	156						
Variable		Mean	Std Dev				
Healthy Life Expectancy		0.597346	0.247579				
Happiness S	core	5.375917	1.119506				

## Linear Fit

Happiness Score = 3.2803744 + 3.508087\*Healthy Life Expectancy

 Table 10 Summary of Fit (Score vs Health)

Summary of Fit				
RSquare	0.601887			
RSquare Adj	0.599302			
Root Mean Square Error	0.708656			
Mean of Response	5.375917			
Observations (or Sum Wgts)	156			

## Lack of Fit

Table 11 Analysis of Variance (Score vs Health)

Analysis of Variance						
Source	DF	Sum of Squares	Mean Square	F Ratio		
Model	1	116.92280	116.923	232.8246		
Error	154	77.33767	0.502	Prob > F		
C. Total	155	194.26047		<.0001*		

Table 12 Parameter Estimates (Score vs Health)

Parameter Estimates							
Term Estimate Std Error t Ratio Prob> t							
Intercept	3.2803744	0.148594	22.08	<.0001*			
Healthy Life Expectancy	3.508087	0.229909	15.26	<.0001*			

## 3.1.4 Happiness Score vs Freedom

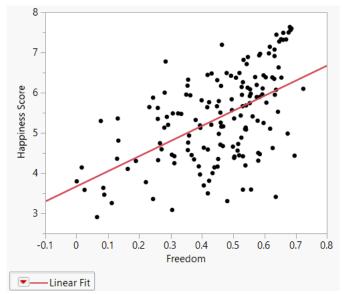


Figure 4 Happiness Score vs Freedom (Scatter Plot)

Table 13 Summary Statistics (Score vs Freedom)

Summary Statistics						
	V	alue	Low	er 95%	Upper 95%	Signif. Prob
Correlation	0.5	4428	0.4	423347	0.646156	<.0001*
Covariance	0.09	8969				
Count		156				
Variable		N	lean	Std De	ev	
Freedom		0.45	4506	0.16242	24	
Happiness S	core	5.37	5917	1.11950	06	

## Linear Fit

Happiness Score = 3.6708645 + 3.7514369\*Freedom

Table 14 Summary of Fit (Score vs Freedom)

Summary of Fit				
RSquare	0.296241			
RSquare Adj	0.291671			
Root Mean Square Error	0.942202			
Mean of Response	5.375917			
Observations (or Sum Wgts)	156			

### Lack of Fit

 Table 15 Analysis of Variance (Score vs Freedom)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	
Model	1	57.54785	57.5478	64.8248	
Error	154	136.71262	0.8877	Prob > F	
C. Total	155	194.26047		<.0001*	

Table 16 Parameter Estimates (Score vs Freedom)

Parameter Estimates							
Term	Estimate	Std Error	t Ratio	Prob> t			
Intercept	3.6708645	0.224806	16.33	<.0001*			
Freedom	3.7514369	0.465937	8.05	<.0001*			

### 3.1.5 Happiness Score vs Generosity

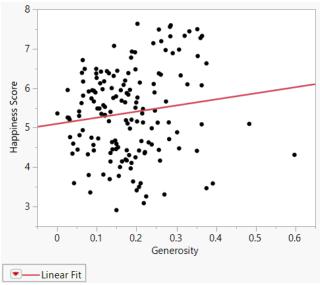


Figure 5 Happiness Score vs Generosity (Scatter Plot)

 Table 17 Summary Statistics (Score vs Generosity)

Summary Statistics						
	٧	alue	Low	er 95%	Upper 95%	Signif. Prob
Correlation	0.13	5825	-(	0.02178	0.286843	0.0909
Covariance	0.01	4973				
Count		156				
Variable		N	/lean	Std De	ev	
Generosity		0.18	1006	0.09847	71	
Happiness S	core	5.37	5917	1.11950	)6	

### Linear Fit

Happiness Score = 5.0964097 + 1.5441828\*Generosity

 Table 18 Summary of Fit (Score vs Generosity)

Summary of Fit				
RSquare	0.018448			
RSquare Adj	0.012075			
Root Mean Square Error	1.112727			
Mean of Response	5.375917			
Observations (or Sum Wgts)	156			

## Lack of Fit

Table 19 Analysis of Variance (Score vs Generosity)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	
Model	1	3.58379	3.58379	2.8944	
Error	154	190.67668	1.23816	Prob > F	
C. Total	155	194.26047		0.0909	

Table 20 Parameter Estimates (Score vs Generosity)

Parameter Estimates						
Term	Estimate	Std Error	t Ratio	Prob> t		
Intercept	5.0964097	0.18689	27.27	<.0001*		
Generosity	1.5441828	0.907645	1.70	0.0909		

## 3.1.6 Happiness Score vs Corruption

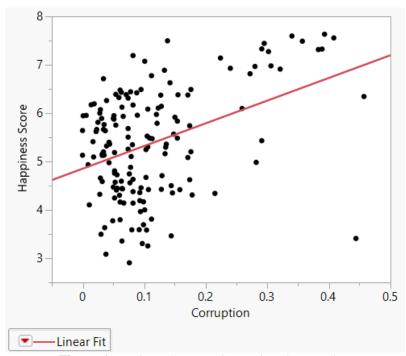


Figure 6 Happiness Score vs Corruption (Scatter Plot)

 Table 21 Summary Statistics (Score vs Corruption)

Summary Statistics							
	V	alue	Low	er 95%	U	pper 95%	Signif. Prob
Correlation	0.40	3234	0.	262742		0.526983	<.0001*
Covariance	0.04	3418					
Count		156					
Variable		N	/lean	Std De	ev		
Corruption		(	0.112	0.0961	18		
Happiness S	core	5.37	5917	1.11950	)6		

## Linear Fit

Happiness Score = 4.8502422 + 4.693522\*Corruption

Table 22 Summary of Fit (Score vs Corruption)

Summary of Fit			
RSquare	0.162598		
RSquare Adj	0.15716		
Root Mean Square Error	1.027777		
Mean of Response	5.375917		
Observations (or Sum Wqts)	156		

#### Lack of Fit

Table 23 Analysis of Variance (Score vs Corruption)

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Ratio	
Model	1	31.58632	31.5863	29.9021	
Error	154	162.67415	1.0563	Prob > F	
C. Total	155	194.26047		<.0001*	

 Table 24 Parameter Estimates (Score vs Corruption)

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	4.8502422	0.126541	38.33	<.0001*
Corruption	4.693522	0.858318	5.47	<.0001*

### 3.2 Model Analysis

In the model analysis, the standard least-squares method is used to predict the regression equation. The following procedure is followed:

#### 3.2.1 Check for Multicollinearity

To check for multicollinearity, the model is analyzed with all the regressors and the variance inflation factor (VIF) is computed for every regressor. VIF is a measure of how much the standard error of the estimate of the coefficient is inflated due to multicollinearity. Following results are obtained from the JMP:

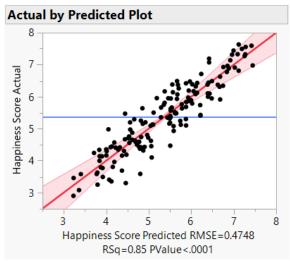
 Table 25 Parameter Estimates (Multicollinearity)

Parameter Estimates					
Term	Estimate	Std Error	t Ratio	Prob> t	VIF
Intercept	1.8849523	0.194758	9.68	<.0001*	
GDP per capita	1.0942376	0.210199	5.21	<.0001*	3.8269757
Social Support	1.009819	0.201858	5.00	<.0001*	2.1007494
Healthy Life Expectancy	0.8150497	0.330185	2.47	0.0147*	3.7682481
Freedom	1.3687377	0.319551	4.28	<.0001*	1.5190808
Generosity	0.5798965	0.47271	1.23	0.2219	1.2218032
Corruption	0.6823737	0.529596	1.29	0.1996	1.4630467

The value of VIFs is well under 5 which shows there is no collinearity (or multicollinearity) in the regressors.

#### 3.2.2 Full-Model Analysis

For all the data points (156 data sets), the two-level interaction is considered (square terms and individual factor interactions) and the model is analyzed. Following results are obtained:



**Figure 7** Actual by Predicted Plot (Full-Model)

Table 26 Effect Summary (Full-Model Analysis)

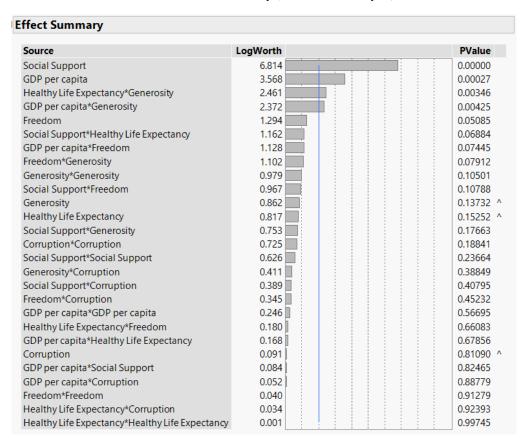


Table 27 Summary of Fit (Full-Model Analysis)

Summary of Fit		
RSquare	0.851431	
RSquare Adj	0.820092	
Root Mean Square Error	0.474844	
Mean of Response	5.375917	
Observations (or Sum Wgts)	156	

Table 28 Analysis of Variance (Full-Model Analysis)

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	27	165.39938	6.12590	27.1686
Error	128	28.86109	0.22548	Prob > F
C. Total	155	194.26047		<.0001*

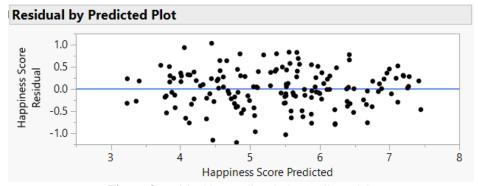


Figure 8 Residual by Predicted Plot (Full-Model)

It can be observed from the residual plot that there is no pattern in the plot. All the points are distributed properly about the zero value.

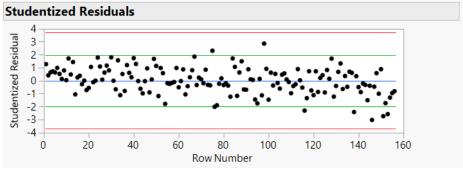


Figure 9 Studentized Residual Plot (Full-Model)

From figure 9, it can be observed that there is no outlier in the data set.

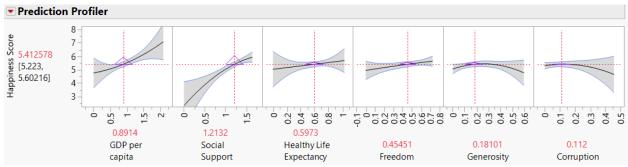


Figure 10 Prediction Profiler (Full-Model)

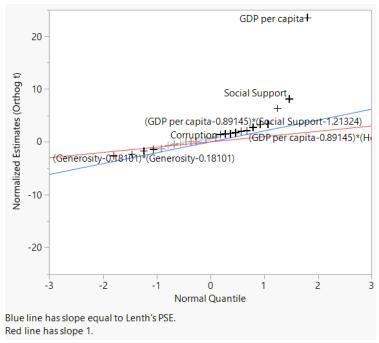


Figure 11 Normal Plot (Full-Model)

**Table 29** Parameter Estimates (Full-Model Analysis)

Parameter Estimates				
Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	1.6244498	0.317265	5.12	<.0001*
GDP per capita	1.0089269	0.269308	3.75	0.0003*
Social Support	1.5747722	0.283462	5.56	<.0001*
Healthy Life Expectancy	0.6289583	0.437003	1.44	0.1525
Freedom	0.9191695	0.466278	1.97	0.0508
Generosity	0.8898182	0.595106	1.50	0.1373
Corruption	0.2119885	0.884151	0.24	0.8109
(GDP per capita-0.89145)*(Social Support-1.21324)	0.2511561	1.131188	0.22	0.8246
(GDP per capita-0.89145)*(Healthy Life Expectancy-0.59735)	-0.964234	2.321341	-0.42	0.6786
(GDP per capita-0.89145)*(Freedom-0.45451)	-3.85382	2.142769	-1.80	0.0745
(GDP per capita-0.89145)*(Generosity-0.18101)	9.6452751	3.313075	2.91	0.0042*
(GDP per capita-0.89145)*(Corruption-0.112)	-0.610144	4.315657	-0.14	0.8878
(Social Support-1.21324)*(Healthy Life Expectancy-0.59735)	3.0912197	1.684669	1.83	0.0688
(Social Support-1.21324)*(Freedom-0.45451)	2.5703252	1.587464	1.62	0.1079
(Social Support-1.21324)*(Generosity-0.18101)	3.5503653	2.613043	1.36	0.1766
(Social Support-1.21324)*(Corruption-0.112)	2.9260903	3.524376	0.83	0.4079
(Healthy Life Expectancy-0.59735)*(Freedom-0.45451)	1.1894092	2.704495	0.44	0.6608
(Healthy Life Expectancy-0.59735)*(Generosity-0.18101)	-13.61276	4.569412	-2.98	0.0035*
(Healthy Life Expectancy-0.59735)*(Corruption-0.112)	0.6849222	7.159366	0.10	0.9239
(Freedom-0.45451)*(Generosity-0.18101)	-7.338582	4.146282	-1.77	0.0791
(Freedom-0.45451)*(Corruption-0.112)	5.3688327	7.121755	0.75	0.4523
(Generosity-0.18101)*(Corruption-0.112)	5.6736082	6.556717	0.87	0.3885
(GDP per capita-0.89145)*(GDP per capita-0.89145)	0.31575	0.550049	0.57	0.5669
(Social Support-1.21324)*(Social Support-1.21324)	-0.797915	0.671075	-1.19	0.2366
(Healthy Life Expectancy-0.59735)*(Healthy Life Expectancy-0.59735)	0.0064845	2.028107	0.00	0.9975
(Freedom-0.45451)*(Freedom-0.45451)	-0.203873	1.857852	-0.11	0.9128
(Generosity-0.18101)*(Generosity-0.18101)	-5.809434	3.558424	-1.63	0.1050
(Corruption-0.112)*(Corruption-0.112)	-6.977637	5.276729	-1.32	0.1884

From the effect summary (table 26), none of the square terms is significant in this model and only 8 terms (considering individual regressors and 2 interaction terms with a p-value less than 5 %) are significant. So, the next step is to analyze the reduced model with only significant terms and get the prediction equation which can be later used for the model validation.

#### 3.2.3 Reduced-Model Analysis

The terms which are considered in the reduced model analysis are: GDP per capita, Social Support, Health Life Expectancy, Freedom, Generosity, Corruption, Health Life Expectancy\*Generosity, and GDP per capita\*Generosity.

Following results are obtained for the reduced model –

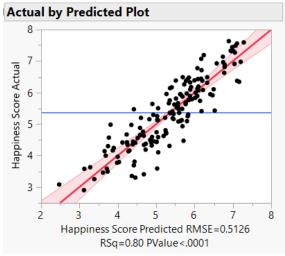


Figure 12 Actual by Predicted Plot (Reduced-Model)

Table 30 Effect Summary (Reduced-Model Analysis)

_		
Source	LogWorth	PValue
Social Support	6.484	0.00000
GDP per capita	6.084	0.00000
Freedom	4.532	0.00003
GDP per capita*Generosity	2.392	0.00405
Healthy Life Expectancy*Generosity	1.511	0.03085
Healthy Life Expectancy	1.199	0.06317
Generosity	0.543	0.28637
Corruption	0.452	0.35301

 Table 31 Summary of Fit (Reduced-Model Analysis)

Summary of Fit				
RSquare	0.801174			
RSquare Adj	0.790354			
Root Mean Square Error	0.51259			
Mean of Response	5.375917			
Observations (or Sum Wgts)	156			

Table 32 Analysis of Variance (Reduced-Model Analysis)

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	8	155.63651	19.4546	74.0427
Error	147	38.62396	0.2627	Prob > F
C. Total	155	194.26047		<.0001*

From the summary of the fit of the reduced model, it can be observed that there is not much significant difference in the RSquare and RSquare Adj value.

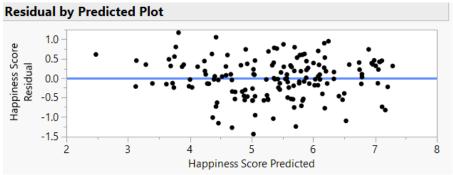


Figure 13 Residual by Predicted Plot (Reduced-Model)

In the residual plot above, all the points are widely spread and there is no pattern in the data points.

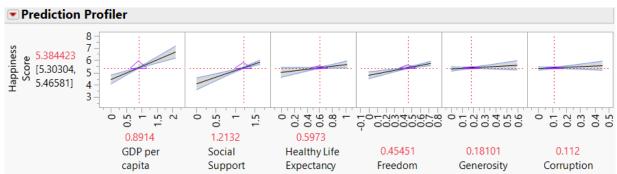


Figure 14 Prediction Profiler (Reduced-Model)

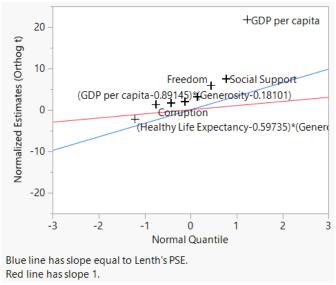


Figure 15 Normal Plot (Reduced-Model)

Therefore, from the analysis above, the prediction equation is given as:

```
Happiness Score
```

```
= 1.9701022982 + (1.0780197168 * GDP per capita)
+ (1.083974302 * Social Support)
+ (0.6342282528 * Health Life Expectancy) + (1.3479943329 * Freedom)
+ (0.4963592909 * Generosity) + (0.5084737571 * Corruption)
+ (GDP per capita - 0.8914487179)
* ((Generosity - 0.1810064103) * 8.6397046272)
+ (Health Life Expectancy - 0.5973461538)
* ((Generosity - 0.1810064103) * (-9.744276207))
```

#### 3.3 Model Adequacy Checking

To check the model adequacy, two plots are considered. One is the studentized residual plot and the second is the distribution of residuals or error in the response variable.

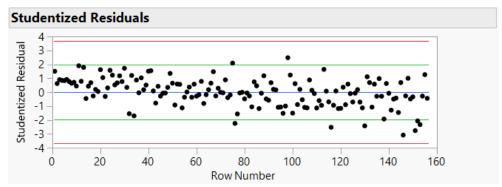


Figure 16 Studentized Residuals (Reduced-Model)

From figure 16, it can be observed that there is no outlier in the model and all the values are under the limits.

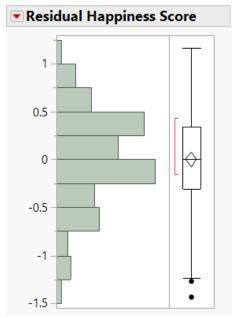


Figure 17 Residual Happiness Score (Model Adequacy)

Table 33 Summary Statistics (Model Adequacy)

<b>▼</b> Summary Statistics		
Mean	-1.08e-16	
Std Dev	0.4991862	
Std Err Mean	0.0399669	
Upper 95% Mean	0.0789501	
Lower 95% Mean	-0.07895	
N	156	

In figure 17, the error is normally distributed, and the mean value is very close to zero (Table 32).

Therefore, it can be concluded that the assumptions in regression analysis are satisfied, and the reduced model is adequate.

#### 3.4 Model Validation

To validate the model, the 156 data points for the year 2019 (<u>View data</u>) are taken and the value of happiness score is predicted for all the data points (using excel) from the regression (prediction) equation obtained above. The table is imported in the JMP and the distribution of regression is plotted.

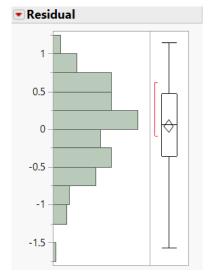


Figure 18 Residual Plot (Model Validation)

 Table 34 Summary Statistics (Model Validation)

<b>▼</b> Summary Statistics			
Mean	0.0374544		
Std Dev	0.5357657		
Std Err Mean	0.0428956		
Upper 95% Mean	0.1221898		
Lower 95% Mean	-0.047281		
N	156		

The residual is normally distributed, and the mean value is approximately equal to zero.

Also, from the new prediction data, the average squared prediction error is calculated using the formula:

$$\frac{\sum_{i=1}^{156} (y_i - \hat{y}_i)^2}{156}$$
$$= 0.286608$$

where,  $y_i$  is the actual value and  $\hat{y}_i$  is the predicted value.

The mean square residual from the fitted model (reduced model) (from table 32),

$$MS_{res} = 0.2627$$

The excel sheet for the above calculations can be found <u>here</u>.

The  $MS_{res}$  is smaller but very close to the average squared prediction error. Therefore, the least square model is a successful predictor for the new data as well as it fits the existing data.

#### 4. Conclusion

In the regression analysis above, the happiness score for 156 countries is considered as a response variable and the six regressors namely GDO per capita, Social Support, Health Life Expectancy, Generosity, Freedom, and Corruption. The response variable is analyzed individually with all the regressors. Then, after that to check the multicollinearity, the model is analyzed with all the regressors together without any interaction terms and the variance inflation factor is computed. It is found that there is no multicollinearity between regressors as the value of VIFs is under 5. After that, the full model is analyzed with all the regressors using the standard least-squares method. From all the 27 terms, 8 terms were considered to analyze the reduced model and get the prediction equation of the analysis. The model adequacy test is performed and the studentized residual graph and the distribution of residuals are plotted. There is no outlier and the residuals are normally distributed. To validate the data, model validation is performed and do that, the data set (156 points) for the year 2019 is considered and the value of response (happiness score) is calculated. It is concluded that the value of the residual mean square of the fitted model (reduced model) and the averaged squared prediction error from the new prediction data is almost the same which validates the prediction equation.

## 5. References

https://www.kaggle.com/unsdsn/world-happiness

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2013). *Introduction to linear regression analysis*. Oxford: Wiley-Blackwell.

Montgomery, D. C. (2020). Design and analysis of experiments. Hoboken, NJ: Wiley.

The excel sheets for the data set can be found here:

 $\frac{https://docs.google.com/spreadsheets/d/1g \ 6nXti0BpWBW8MpE1BnMgYcFJTtNKQxgLxK5lP \ 1V6k/edit?usp=sharing}{}$ 

 $\underline{https://docs.google.com/spreadsheets/d/1Ek4\_CTUSrsRAAkZBR7wkNpI1IrPYdoTTLQFQ-lEbJPY/edit?usp=sharing}$ 

## Data Set for the year 2018

Happiness	GDP per	Social	Healthy Life			
Score	capita	Support	Expectancy	Freedom	Generosity	Corruption
7.632	1.305	1.592	0.874	0.681	0.202	0.393
7.594	1.456	1.582	0.861	0.686	0.286	0.34
7.555	1.351	1.59	0.868	0.683	0.284	0.408
7.495	1.343	1.644	0.914	0.677	0.353	0.138
7.487	1.42	1.549	0.927	0.66	0.256	0.357
7.441	1.361	1.488	0.878	0.638	0.333	0.295
7.328	1.33	1.532	0.896	0.653	0.321	0.291
7.324	1.268	1.601	0.876	0.669	0.365	0.389
7.314	1.355	1.501	0.913	0.659	0.285	0.383
7.272	1.34	1.573	0.91	0.647	0.361	0.302
7.19	1.244	1.433	0.888	0.464	0.262	0.082
7.139	1.341	1.504	0.891	0.617	0.242	0.224
7.072	1.01	1.459	0.817	0.632	0.143	0.101
6.977	1.448	1.583	0.876	0.614	0.307	0.306
6.965	1.34	1.474	0.861	0.586	0.273	0.28
6.927	1.324	1.483	0.894	0.583	0.188	0.24
6.91	1.576	1.52	0.896	0.632	0.196	0.321
6.886	1.398	1.471	0.819	0.547	0.291	0.133
6.814	1.301	1.559	0.883	0.533	0.354	0.272
6.774	2.096	0.776	0.67	0.284	0.186	0.112
6.711	1.233	1.489	0.854	0.543	0.064	0.034
6.627	1.27	1.525	0.884	0.645	0.376	0.142
6.489	1.293	1.466	0.908	0.52	0.098	0.176
6.488	1.038	1.252	0.761	0.479	0.069	0.095
6.476	1.131	1.331	0.808	0.431	0.197	0.061
6.441	1.365	1.436	0.857	0.418	0.151	0.078
6.43	1.112	1.438	0.759	0.597	0.125	0.063
6.419	0.986	1.474	0.675	0.493	0.11	0.088
6.388	1.073	1.468	0.744	0.57	0.062	0.054
6.382	0.781	1.268	0.608	0.604	0.179	0.071
6.379	1.093	1.459	0.771	0.625	0.13	0.155
6.374	1.649	1.303	0.748	0.654	0.256	0.171
6.371	1.379	1.331	0.633	0.509	0.098	0.127
6.343	1.529	1.451	1.008	0.631	0.261	0.457
6.322	1.161	1.258	0.669	0.356	0.311	0.059
6.31	1.251	1.538	0.965	0.449	0.142	0.074
6.26	0.96	1.439	0.635	0.531	0.099	0.039
6.192	1.223	1.492	0.564	0.575	0.171	0.019
6.173	1.21	1.537	0.776	0.354	0.118	0.014

6.167	0.806	1.231	0.639	0.461	0.065	0.082
6.141	0.668	1.319	0.7	0.527	0.208	0.128
6.123	1.176	1.448	0.781	0.546	0.108	0.064
6.105	1.338	1.366	0.698	0.594	0.243	0.123
6.096	0.719	1.584	0.605	0.724	0.328	0.259
6.083	1.474	1.301	0.675	0.554	0.167	0.106
6.072	1.016	1.417	0.707	0.637	0.364	0.029
6	1.264	1.501	0.946	0.281	0.137	0.028
5.973	0.889	1.33	0.736	0.556	0.114	0.12
5.956	0.807	1.101	0.474	0.593	0.183	0.089
5.952	1.197	1.527	0.716	0.35	0.026	0.006
5.948	1.219	1.506	0.856	0.633	0.16	0.051
5.945	1.116	1.219	0.726	0.528	0.088	0.001
5.933	1.148	1.454	0.671	0.363	0.092	0.066
5.915	1.294	1.462	0.988	0.553	0.079	0.15
5.891	1.09	1.387	0.684	0.584	0.245	0.05
5.89	0.819	1.493	0.693	0.575	0.096	0.031
5.875	1.266	1.204	0.955	0.244	0.175	0.051
5.835	1.229	1.211	0.909	0.495	0.179	0.154
5.81	1.151	1.479	0.599	0.399	0.065	0.025
5.79	1.143	1.516	0.631	0.454	0.148	0.121
5.762	1.229	1.191	0.909	0.423	0.202	0.035
5.752	0.751	1.223	0.508	0.606	0.141	0.054
5.739	1.2	1.532	0.737	0.553	0.086	0.174
5.681	0.835	1.522	0.615	0.541	0.162	0.074
5.663	0.934	1.249	0.674	0.53	0.092	0.034
5.662	0.855	1.23	0.578	0.448	0.274	0.023
5.64	0.657	1.301	0.62	0.232	0.171	0
5.636	1.016	1.533	0.517	0.417	0.199	0.037
5.62	1.171	1.401	0.732	0.259	0.061	0.022
5.566	0.985	1.35	0.553	0.496	0.116	0.148
5.524	0.775	1.312	0.513	0.643	0.12	0.105
5.504	0.62	1.205	0.622	0.459	0.197	0.074
5.483	1.039	1.498	0.7	0.307	0.101	0.154
5.483	1.148	1.38	0.686	0.324	0.106	0.109
5.472	0.652	0.81	0.424	0.334	0.216	0.113
5.43	1.405	1.29	1.03	0.524	0.246	0.291
5.41	1.188	1.429	0.884	0.562	0.055	0.017
5.398	0.975	1.369	0.685	0.288	0.134	0.043
5.358	1.154	1.202	0.879	0.131	0	0.044
5.358	0.965	1.179	0.785	0.503	0.214	0.136
5.347	1.017	1.279	0.729	0.259	0.111	0.081
5.321	1.115	1.161	0.737	0.38	0.12	0.039

E 2/						
5.30		1.441	0.614	0.578	0.12	0.106
5.29		1.154	0.687	0.077	0.055	0.135
5.25		0.797	0.669	0.46	0.026	0.074
5.24	6 0.989	1.142	0.799	0.597	0.029	0.103
5.20		1.161	0.603	0.43	0.031	0.176
5.19	9 0.474	1.166	0.598	0.292	0.187	0.034
5.18	0.959	1.239	0.691	0.394	0.173	0.052
5.16	0.822	1.265	0.645	0.468	0.13	0.134
5.15	0.689	1.172	0.048	0.462	0.201	0.032
5.13	0.53	1.416	0.594	0.54	0.281	0.035
5.12	9 0.915	1.078	0.758	0.28	0.216	0
5.12	5 0.914	1.517	0.575	0.395	0.253	0.032
5.10	0.715	1.365	0.702	0.618	0.177	0.079
5.09	0.899	1.215	0.522	0.538	0.484	0.018
5.08	0.796	1.335	0.527	0.541	0.364	0.171
4.98	2 0	0.712	0.115	0.674	0.238	0.282
4.97	0.535	0.891	0.182	0.454	0.183	0.043
4.93	1.054	1.515	0.712	0.359	0.064	0.009
4.8	0.425	1.228	0.539	0.526	0.302	0.078
4.80	0.996	1.469	0.657	0.133	0.056	0.052
4.75	1.036	1.164	0.404	0.356	0.032	0.052
4.74	3 0.642	1.217	0.602	0.266	0.086	0.076
4.72	0.94	1.41	0.33	0.516	0.103	0.056
4.70	1.059	0.771	0.691	0.459	0.282	0.129
4.67	0.541	0.872	0.08	0.467	0.146	0.103
4.65	7 0.592	0.896	0.337	0.499	0.212	0.029
4.63	0.429	1.117	0.433	0.406	0.138	0.082
4.62	3 0.72	1.034	0.441	0.626	0.23	0.174
4.59	0.9	0.906	0.69	0.271	0.04	0.063
4.58	6 0.916	0.817	0.79	0.419	0.149	0.032
4.57	0.256	0.813	0	0.355	0.238	0.053
4.55	9 0.682	0.811	0.343	0.514	0.091	0.077
4	5 0.532	0.85	0.579	0.58	0.153	0.144
4.47	0.918	1.314	0.672	0.585	0.307	0.05
4.45	6 1.01	0.971	0.536	0.304	0.148	0.095
4.44	7 0.37	1.233	0.152	0.367	0.139	0.056
4.44	1 0.874	1.281	0.365	0.519	0.051	0.064
4.43	3 0.549	1.088	0.457	0.696	0.256	0.065
4.42	4 0.314	1.097	0.254	0.312	0.175	0.128
4.41	9 0.885	1.025	0.553	0.312	0.092	0.107
4.41	7 0.198	0.902	0.173	0.531	0.206	0.158
4.4	1 0.493	1.048	0.454	0.504	0.352	0.055
4.37		1.047	0.295	0.503	0.221	0.082

4.356	0.557	1.245	0.292	0.129	0.134	0.093
4.35	0.308	0.95	0.391	0.452	0.22	0.146
4.34	0.853	0.592	0.643	0.375	0.038	0.215
4.321	0.816	0.99	0.666	0.26	0.077	0.028
4.308	0.682	1.174	0.429	0.58	0.598	0.178
4.301	0.358	0.907	0.053	0.189	0.181	0.06
4.245	0.069	1.136	0.204	0.312	0.197	0.052
4.19	0.721	0.747	0.485	0.539	0.172	0.093
4.166	0.131	0.867	0.221	0.39	0.175	0.099
4.161	0.322	1.09	0.237	0.45	0.259	0.061
4.141	0.378	0.372	0.24	0.44	0.163	0.067
4.139	0.605	1.24	0.312	0.016	0.134	0.082
4.103	0.793	1.413	0.609	0.163	0.187	0.011
3.999	0.259	0.474	0.253	0.434	0.158	0.101
3.964	0.344	0.792	0.211	0.394	0.185	0.094
3.808	0.472	1.215	0.079	0.423	0.116	0.112
3.795	0.73	1.125	0.269	0	0.079	0.061
3.774	0.262	0.908	0.402	0.221	0.155	0.049
3.692	0.357	1.094	0.248	0.406	0.132	0.099
3.632	0.332	0.537	0.255	0.085	0.191	0.036
3.59	1.017	1.174	0.417	0.557	0.042	0.092
3.587	0.186	0.541	0.306	0.531	0.21	0.08
3.582	0.315	0.714	0.289	0.025	0.392	0.104
3.495	0.076	0.858	0.267	0.419	0.206	0.03
3.462	0.689	0.382	0.539	0.088	0.376	0.144
3.408	0.332	0.896	0.4	0.636	0.2	0.444
3.355	0.442	1.073	0.343	0.244	0.083	0.064
3.303	0.455	0.991	0.381	0.481	0.27	0.097
3.254	0.337	0.608	0.177	0.112	0.224	0.106
3.083	0.024	0	0.01	0.305	0.218	0.038
2.905	0.091	0.627	0.145	0.065	0.149	0.076

## Data Set for the year 2019 (Validation Data)

Happiness	GDP per	Social	Healthy Life			
Score	capita	Support	Expectancy	Freedom	Generosity	Corruption
7.769	1.34	1.587	0.986	0.596	0.153	0.393
7.6	1.383	1.573	0.996	0.592	0.252	0.41
7.554	1.488	1.582	1.028	0.603	0.271	0.341
7.494	1.38	1.624	1.026	0.591	0.354	0.118
7.488	1.396	1.522	0.999	0.557	0.322	0.298
7.48	1.452	1.526	1.052	0.572	0.263	0.343
7.343	1.387	1.487	1.009	0.574	0.267	0.373
7.307	1.303	1.557	1.026	0.585	0.33	0.38
7.278	1.365	1.505	1.039	0.584	0.285	0.308
7.246	1.376	1.475	1.016	0.532	0.244	0.226
7.228	1.372	1.548	1.036	0.557	0.332	0.29
7.167	1.034	1.441	0.963	0.558	0.144	0.093
7.139	1.276	1.455	1.029	0.371	0.261	0.082
7.09	1.609	1.479	1.012	0.526	0.194	0.316
7.054	1.333	1.538	0.996	0.45	0.348	0.278
7.021	1.499	1.553	0.999	0.516	0.298	0.31
6.985	1.373	1.454	0.987	0.495	0.261	0.265
6.923	1.356	1.504	0.986	0.473	0.16	0.21
6.892	1.433	1.457	0.874	0.454	0.28	0.128
6.852	1.269	1.487	0.92	0.457	0.046	0.036
6.825	1.503	1.31	0.825	0.598	0.262	0.182
6.726	1.3	1.52	0.999	0.564	0.375	0.151
6.595	1.07	1.323	0.861	0.433	0.074	0.073
6.592	1.324	1.472	1.045	0.436	0.111	0.183
6.446	1.368	1.43	0.914	0.351	0.242	0.097
6.444	1.159	1.369	0.92	0.357	0.187	0.056
6.436	0.8	1.269	0.746	0.535	0.175	0.078
6.375	1.403	1.357	0.795	0.439	0.08	0.132
6.374	1.684	1.313	0.871	0.555	0.22	0.167
6.354	1.286	1.484	1.062	0.362	0.153	0.079
6.321	1.149	1.442	0.91	0.516	0.109	0.054
6.3	1.004	1.439	0.802	0.39	0.099	0.086
6.293	1.124	1.465	0.891	0.523	0.127	0.15
6.262	1.572	1.463	1.141	0.556	0.271	0.453
6.253	0.794	1.242	0.789	0.43	0.093	0.074
6.223	1.294	1.488	1.039	0.231	0.158	0.03
6.199	1.362	1.368	0.871	0.536	0.255	0.11
6.198	1.246	1.504	0.881	0.334	0.121	0.014
6.192	1.231	1.477	0.713	0.489	0.185	0.016

6.182	1.206	1.438	0.884	0.483	0.117	0.05
6.174	0.745	1.529	0.756	0.631	0.322	0.24
6.149	1.238	1.515	0.818	0.291	0.043	0.042
6.125	0.985	1.41	0.841	0.47	0.099	0.034
6.118	1.258	1.523	0.953	0.564	0.144	0.057
6.105	0.694	1.325	0.835	0.435	0.2	0.127
6.1	0.882	1.232	0.758	0.489	0.262	0.006
6.086	1.092	1.432	0.881	0.471	0.066	0.05
6.07	1.162	1.232	0.825	0.462	0.083	0.005
6.046	1.263	1.223	1.042	0.406	0.19	0.041
6.028	0.912	1.312	0.868	0.498	0.126	0.087
6.021	1.5	1.319	0.808	0.493	0.142	0.097
6.008	1.05	1.409	0.828	0.557	0.359	0.028
5.94	1.187	1.465	0.812	0.264	0.075	0.064
5.895	1.301	1.219	1.036	0.159	0.175	0.056
5.893	1.237	1.528	0.874	0.495	0.103	0.161
5.89	0.831	1.478	0.831	0.49	0.107	0.028
5.888	1.12	1.402	0.798	0.498	0.215	0.06
5.886	1.327	1.419	1.088	0.445	0.069	0.14
5.86	0.642	1.236	0.828	0.507	0.246	0.078
5.809	1.173	1.508	0.729	0.41	0.146	0.096
5.779	0.776	1.209	0.706	0.511	0.137	0.064
5.758	1.201	1.41	0.828	0.199	0.081	0.02
5.743	0.855	1.475	0.777	0.514	0.184	0.08
5.718	1.263	1.252	1.042	0.417	0.191	0.162
5.697	0.96	1.274	0.854	0.455	0.083	0.027
5.693	1.221	1.431	0.999	0.508	0.047	0.025
5.653	0.677	0.886	0.535	0.313	0.22	0.098
5.648	1.183	1.452	0.726	0.334	0.082	0.031
5.631	0.807	1.293	0.657	0.558	0.117	0.107
5.603	1.004	1.383	0.854	0.282	0.137	0.039
5.529	0.685	1.328	0.739	0.245	0.181	0
5.525	1.044	1.303	0.673	0.416	0.133	0.152
5.523	1.051	1.361	0.871	0.197	0.142	0.08
5.467	0.493	1.098	0.718	0.389	0.23	0.144
5.432	1.155	1.266	0.914	0.296	0.119	0.022
5.43	1.438	1.277	1.122	0.44	0.258	0.287
5.425	1.015	1.401	0.779	0.497	0.113	0.101
5.386	0.945	1.212	0.845	0.212	0.263	0.006
5.373	1.183	1.36	0.808	0.195	0.083	0.106
5.339	1.221	1.171	0.828	0.508	0.26	0.024
5.323	1.067	1.465	0.789	0.235	0.094	0.142
5.287	1.181	1.156	0.999	0.067	0	0.034

5.285	0.948	1.531	0.667	0.317	0.235	0.038
5.274	0.983	1.294	0.838	0.345	0.185	0.034
5.265	0.696	1.111	0.245	0.426	0.215	0.041
5.261	0.551	1.438	0.723	0.508	0.3	0.023
5.247	1.052	1.538	0.657	0.394	0.244	0.028
5.211	1.002	1.16	0.785	0.086	0.073	0.114
5.208	0.801	0.782	0.782	0.418	0.036	0.076
5.208	1.043	1.147	0.769	0.351	0.035	0.182
5.197	0.987	1.224	0.815	0.216	0.166	0.027
5.192	0.931	1.203	0.66	0.491	0.498	0.028
5.191	1.029	1.125	0.893	0.521	0.058	0.1
5.175	0.741	1.346	0.851	0.543	0.147	0.073
5.082	0.813	1.321	0.604	0.457	0.37	0.167
5.044	0.549	0.91	0.331	0.381	0.187	0.037
5.011	1.092	1.513	0.815	0.311	0.081	0.004
4.996	0.611	0.868	0.486	0.381	0.245	0.04
4.944	0.569	0.808	0.232	0.352	0.154	0.09
4.913	0.446	1.226	0.677	0.439	0.285	0.089
4.906	0.837	1.225	0.815	0.383	0.11	0.13
4.883	0.393	0.437	0.397	0.349	0.175	0.082
4.812	0.673	0.799	0.508	0.372	0.105	0.093
4.799	1.057	1.183	0.571	0.295	0.043	0.055
4.796	0.764	1.03	0.551	0.547	0.266	0.164
4.722	0.96	1.351	0.469	0.389	0.13	0.055
4.719	0.947	0.848	0.874	0.383	0.178	0.027
4.707	0.96	1.427	0.805	0.154	0.064	0.047
4.7	0.574	1.122	0.637	0.609	0.232	0.062
4.696	0.657	1.247	0.672	0.225	0.103	0.066
4.681	0.45	1.134	0.571	0.292	0.153	0.072
4.668	0	0.698	0.268	0.559	0.243	0.27
4.639	0.879	1.313	0.477	0.401	0.07	0.056
4.628	0.138	0.774	0.366	0.318	0.188	0.102
4.587	0.331	1.056	0.38	0.255	0.177	0.113
4.559	0.85	1.055	0.815	0.283	0.095	0.064
4.548	1.1	0.842	0.785	0.305	0.27	0.125
4.534	0.38	0.829	0.375	0.332	0.207	0.086
4.519	0.886	0.666	0.752	0.346	0.043	0.164
4.516	0.308	0.939	0.428	0.382	0.269	0.167
4.509	0.512	0.983	0.581	0.431	0.372	0.053
4.49	0.57	1.167	0.489	0.066	0.106	0.088
4.466	0.204	0.986	0.39	0.494	0.197	0.138
4.461	0.921	1	0.815	0.167	0.059	0.055
4.456	0.562	0.928	0.723	0.527	0.166	0.143

4.437	1.043	0.98	0.574	0.241	0.148	0.089
4.418	0.094	1.125	0.357	0.269	0.212	0.053
4.39	0.385	1.105	0.308	0.327	0.153	0.052
4.374	0.268	0.841	0.242	0.309	0.252	0.045
4.366	0.949	1.265	0.831	0.47	0.244	0.047
4.36	0.71	1.181	0.555	0.525	0.566	0.172
4.35	0.35	0.766	0.192	0.174	0.198	0.078
4.332	0.82	1.39	0.739	0.178	0.187	0.01
4.286	0.336	1.033	0.532	0.344	0.209	0.1
4.212	0.811	1.149	0	0.313	0.074	0.135
4.189	0.332	1.069	0.443	0.356	0.252	0.06
4.166	0.913	1.039	0.644	0.241	0.076	0.067
4.107	0.578	1.058	0.426	0.431	0.247	0.087
4.085	0.275	0.572	0.41	0.293	0.177	0.085
4.015	0.755	0.765	0.588	0.498	0.2	0.085
3.975	0.073	0.922	0.443	0.37	0.233	0.033
3.973	0.274	0.757	0.505	0.142	0.275	0.078
3.933	0.274	0.916	0.555	0.148	0.169	0.041
3.802	0.489	1.169	0.168	0.359	0.107	0.093
3.775	0.046	0.447	0.38	0.22	0.176	0.18
3.663	0.366	1.114	0.433	0.361	0.151	0.089
3.597	0.323	0.688	0.449	0.026	0.419	0.11
3.488	1.041	1.145	0.538	0.455	0.025	0.1
3.462	0.619	0.378	0.44	0.013	0.331	0.141
3.41	0.191	0.56	0.495	0.443	0.218	0.089
3.38	0.287	1.163	0.463	0.143	0.108	0.077
3.334	0.359	0.711	0.614	0.555	0.217	0.411
3.231	0.476	0.885	0.499	0.417	0.276	0.147
3.203	0.35	0.517	0.361	0	0.158	0.025
3.083	0.026	0	0.105	0.225	0.235	0.035
2.853	0.306	0.575	0.295	0.01	0.202	0.091